Application of Python Scripting for Major Work Destination Analysis

Tom Vo, Jung Seo, and Yunkyung Choi

Research & Analysis Department
Southern California Association of Governments

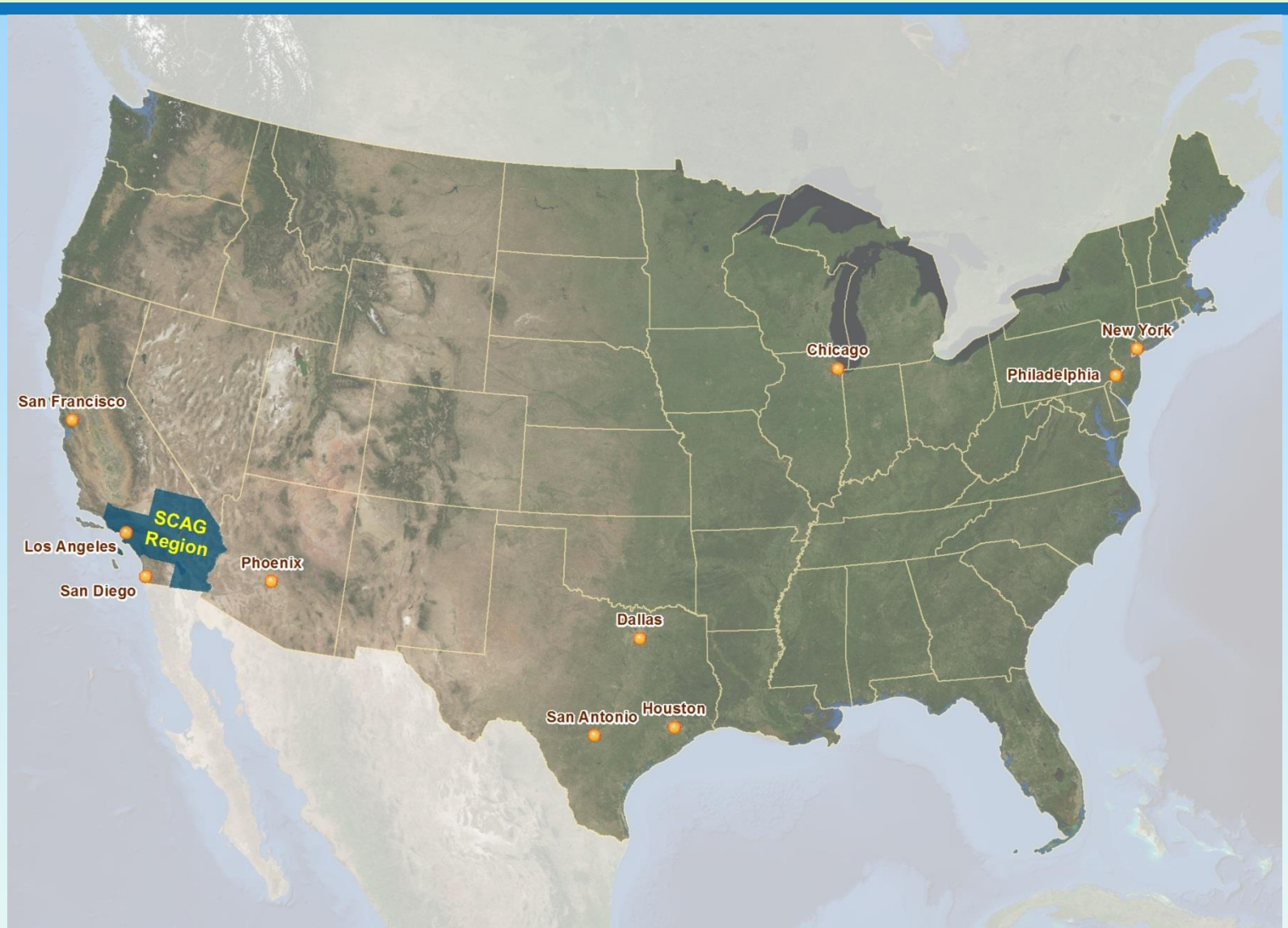
2014 ESRI User Conference | July 17, 2014 | San Diego, CA

SOUTHERN CALIFORNIA ASSOCIATION of GOVERNMENTS

Contents

- Introduction of SCAG
- Objectives
- Methodology
- Results
- Conclusions
- Future Studies

Overview of SCAG



Quick Facts of SCAG

- Nation's largest Metropolitan Planning Organization (MPO)
- 6 counties and 191 cities
- 15 sub-regions
- 18.4 million people (2012)
- 38,000 square miles
- 16th largest economy in the world (GRP: $924 Billion in 2013)

Objectives

- Identify the major work destinations for each jurisdiction in the SCAG region
  - **191 cities and 6 counties**
- Visualize the spatial patterns of the major work destinations for each jurisdiction
  - To understand where residents of each jurisdiction are employed
- Provide informational data resources to local jurisdictions for planning purposes
  - E.g. 2013 Local Profiles: Planning data reports

Methodology for Identifying Work Destination

- Using ***OnTheMap* application**
  - Available on http://onthemap.ces.census.gov/
- Using SAS program with ***LODES* raw datasets**
  - Available on http://lehd.ces.census.gov/data/

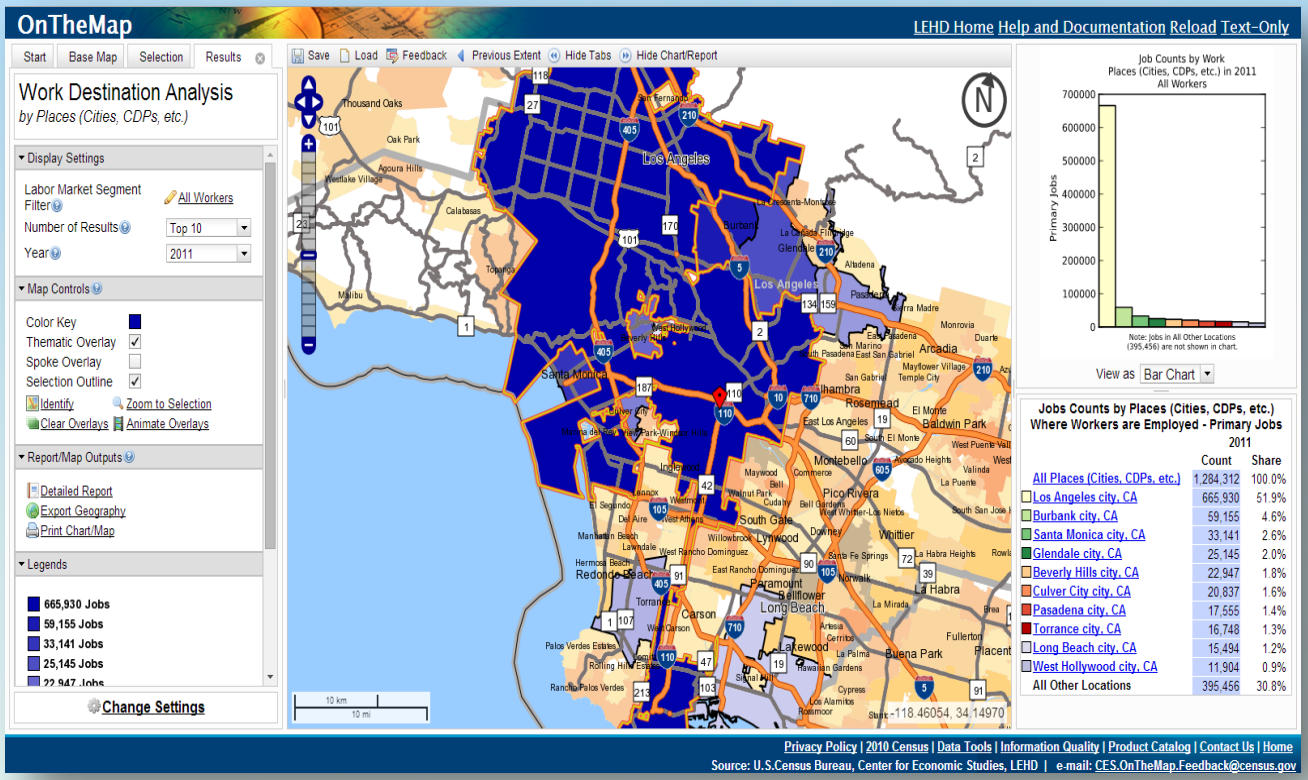
* SAS: Statistical Analysis System

* LODES: LEHD Origin-Destination Employment Statistics

* LEHD: Longitudinal Employer-Household Dynamics

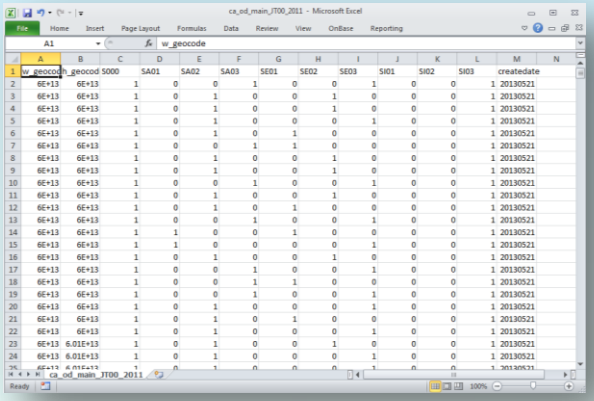
OnTheMap application

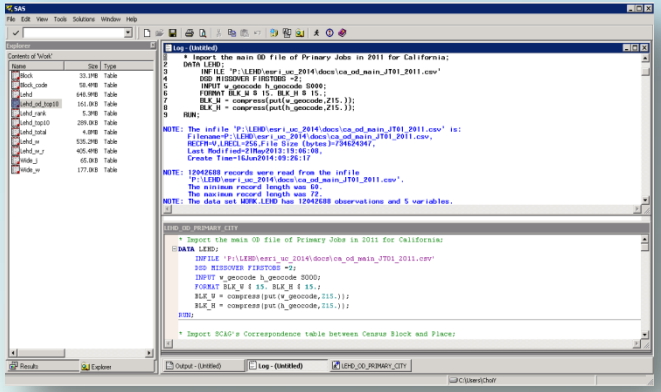
- Web-based mapping and reporting application
- Based on 2002-2011 *LODES* data



Programming Scripts with *LODES* Raw Datasets

- State-based annual data from 2002-2011
- 3 types of CSV data files
  - Origin-Destination (OD)
  - Residence Area Characteristics (RAC)
  - Workplace Area Characteristics (WAC)
- Enumerated by 2010 census blocks





Pros and Cons

	OnTheMap	**Programming Scripts**
Pros	▪ Easy-to-use interface and able to visualize data instantly ▪ Support analyses at multiple levels of geography	▪ Able to manipulate data for multiple areas at once ▪ Improve efficiency of managing and processing big data
Cons	▪ Not able to search/download for multiple locations	▪ Incorporate with ArcGIS to visualize the data

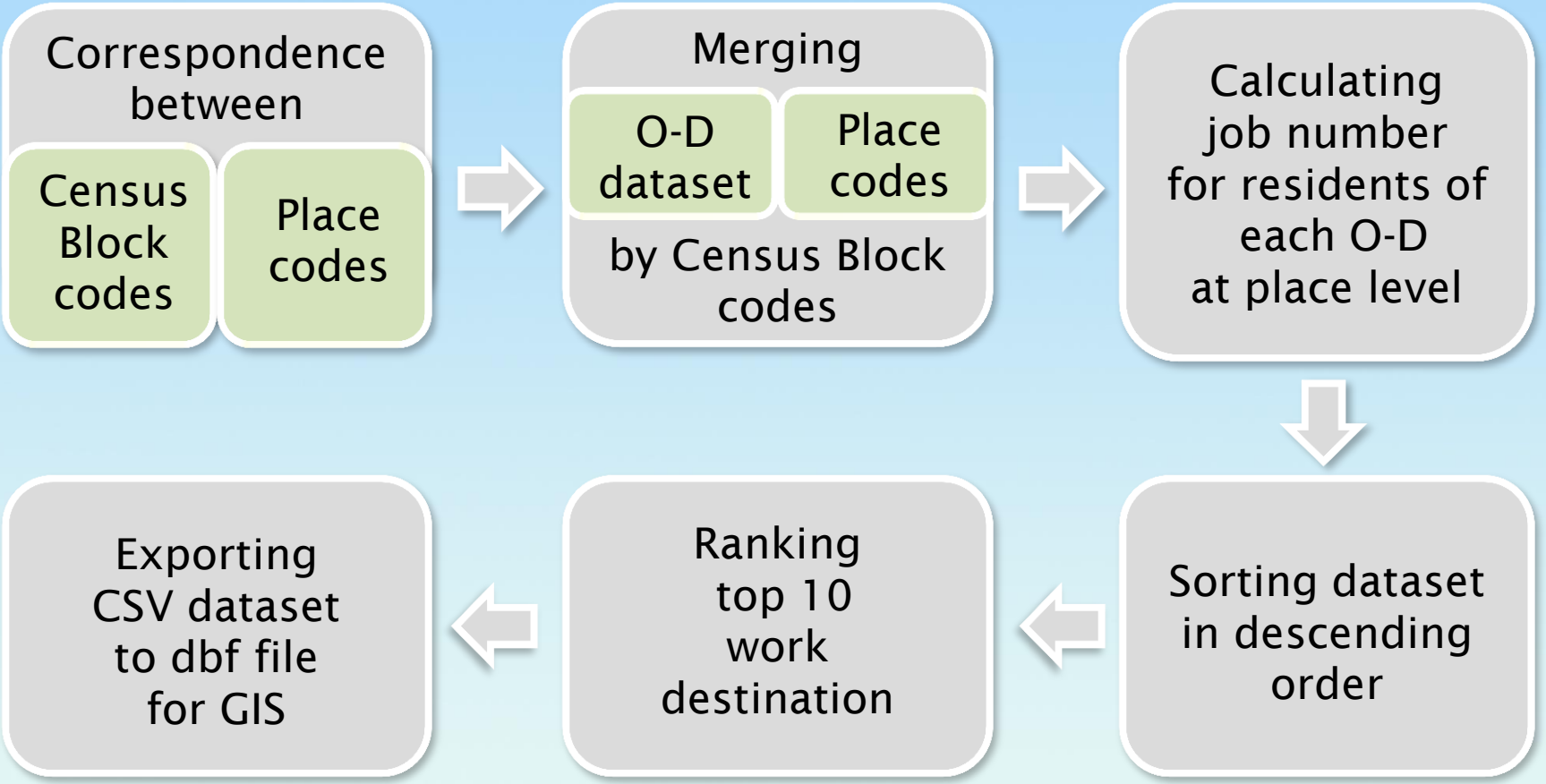
LODES Raw Dataset

- *LODES* raw dataset description
  - *LODES* CA OD dataset (Version 7, 2011)
  - 3 Variables: w_geocode, h_geocode, S000

Pos	Variable	Type	Explanation
1	w_geocode	Char15	Workplace Census Block Code
2	h_geocode	Char15	Residence Census Block Code
3	S000	Num	Total number of jobs
4	SA01	Num	Number of jobs of workers age 29 or younger
5	SA02	Num	Number of jobs of workers age 30 to 54
6	SA03	Num	Number of jobs of workers age 55 or older
7	SE01	Num	Number of jobs with earnings $1250/month or less
8	SE02	Num	Number of jobs with earnings $1251/month to $3333/month
9	SE03	Num	Number of jobs with earnings greater than $3333/month
10	SI01	Num	Number of jobs in Goods Producing industry sectors
11	SI02	Num	Number of jobs in Trade, Transportation, and Utilities industry sectors
12	SI03	Num	Number of jobs in All Other Services industry sectors
13	createdate	Char	Date on which data was created, formatted as YYYYMMDD

Identifying Major Work Destinations

- O-D Analysis Processing Steps



Identifying Major Work Destinations (cont.)

scagmod2 - Remote Desktop Connection

SAS

File Edit View Tools Solutions Window Help

Explorer

Contents of 'Work'

Block, Block_code, Lehd, Lehd_od_to..., Lehd_rank, Lehd_top10, Lehd_total, Lehd_w, Lehd_w_r, Wide_j, Wide_w

Log - (Untitled)

```
120       RETAIN CITY_H CITYNAME_H DES_1 JOB_1 DES_2 JOB_2 DES_3 JOB_3 DES_4 JOB_4 DES_5 JOB_5
120! DES_6 JOB_6 DES_7 JOB_7 DES_8 JOB_8 DES_9 JOB_9 DES_10 JOB_10;
121       SET LEHD_OD_TOP10;
122   RUN;

NOTE: There were 352 observations read from the data set WORK.LEHD_OD_TOP10.
NOTE: The data set WORK.LEHD_OD_TOP10 has 352 observations and 22 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


123
124   * Export data set;
125   PROC EXPORT OUTFILE='P:\LEHD\esri_uc_2014\docs\lehd_od_top10.dbf'
126       DATA=LEHD_OD_TOP10
127       DBMS=DBF REPLACE;
128   RUN;

NOTE: Exported 22 columns and 352 rows of data to the dBase table file.
NOTE: "P:\LEHD\esri_uc_2014\docs\lehd_od_top10.dbf" file was successfully created.
NOTE: PROCEDURE EXPORT used (Total process time):
      real time           0.09 seconds
      cpu time            0.01 seconds
```

LEHD_OD_PRIMARY_CITY

```
* Import the main OD file of Primary Jobs in 2011 for California;
DATA LEHD;
    INFILE 'P:\LEHD\esri_uc_2014\docs\ca_od_main_JT01_2011.csv'
    DSD MISSOVER FIRSTOBS =2;
    INPUT w_geocode h_geocode S000;
    FORMAT BLK_W $ 15. BLK_H $ 15.;
    BLK_W = compress(put(w_geocode,Z15.));
    BLK_H = compress(put(h_geocode,Z15.));
RUN;

* Import SCAG's Correspondence table between Census Block and Place;
PROC IMPORT DATAFILE= 'P:\LEHD\esri_uc_2014\docs\tl_2010_scag_tabblock10_place.dbf'
    OUT=BLOCK_CODE
    DBMS=DBF Replace;
RUN;

* Create Block ID variables for merge purpose;
DATA BLOCK;
    SET BLOCK_CODE;
    FORMAT BLK_W $ 15. BLK_H $ 15.;
```

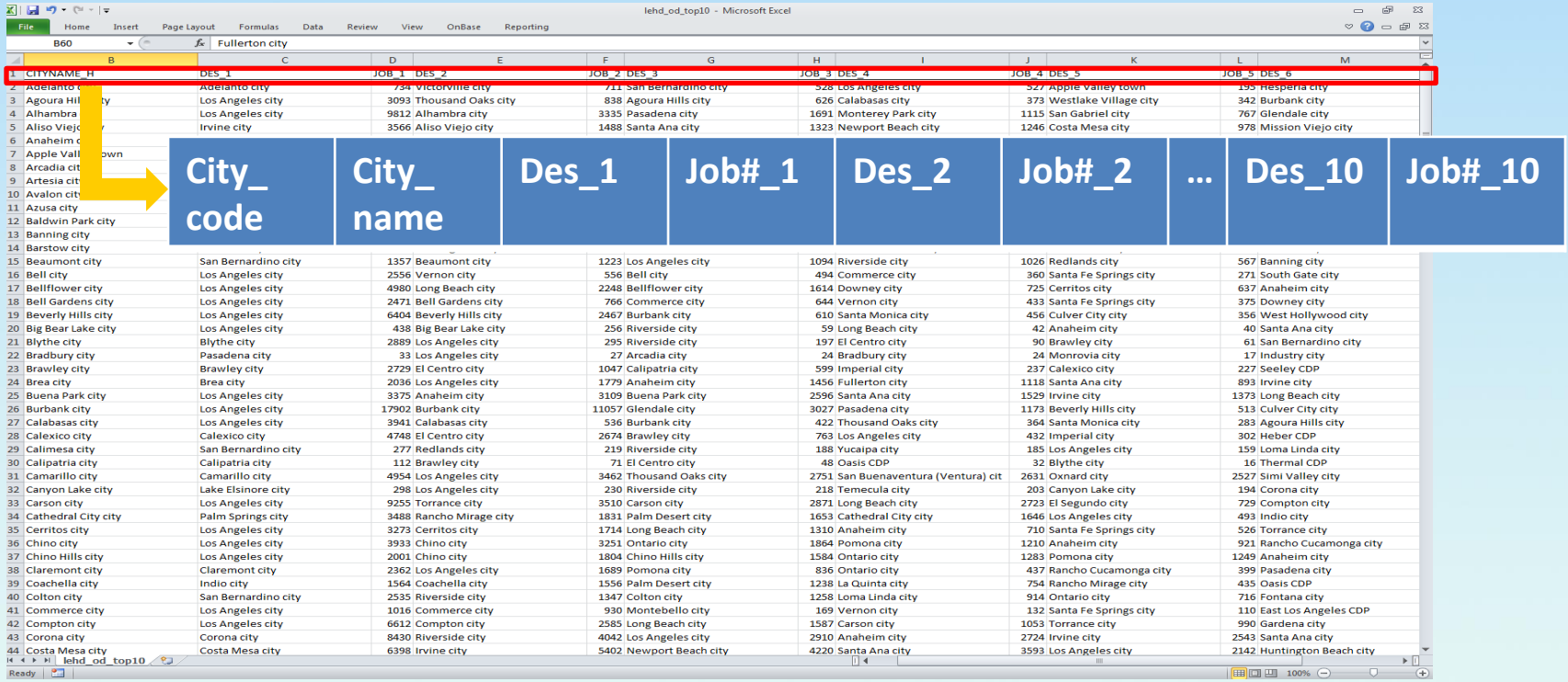
Results | Explorer | Output - (Untitled) | Log - (Untitled) | LEHD_OD_PRIMARY_CITY

C:\Users\ChoiY

Start | SAS | KO

Top 10 Work Destinations Data

- Identified top 10 work destinations where residents of each jurisdictions commute to work (191 cities and 6 counties)



Visualizing Major Work Destinations

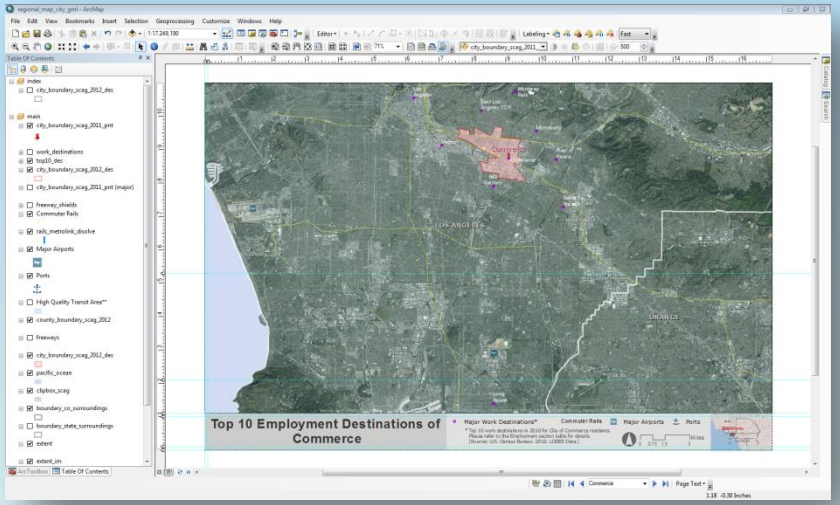
- Using *ESRI ArcGIS* application
  - ***Data Driven Pages*** – To create a multi-page map series from a single map document
- Using *Python* programming language
  - *Python* – Interpreted, object-oriented, high-level general-purpose programming language

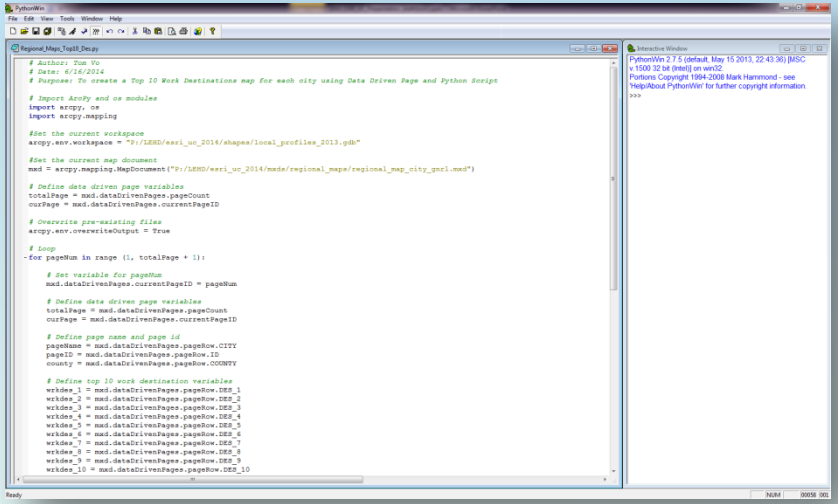
Arcpy Module

- *Python* site package for performing GIS functions available in *ArcGIS*
  - Encompassing the arcgisscripting module
  - E.g. *arcpy.mapping* (mapping module)
- Useful and productive way to perform geographic data analysis, data conversion, data management, and map automation with Python

Visualizing Major Work Destinations
(Automated Mapping Workflow)

- Using ArcGIS and Python to create a series of work destination maps for 191 cities and 6 counties





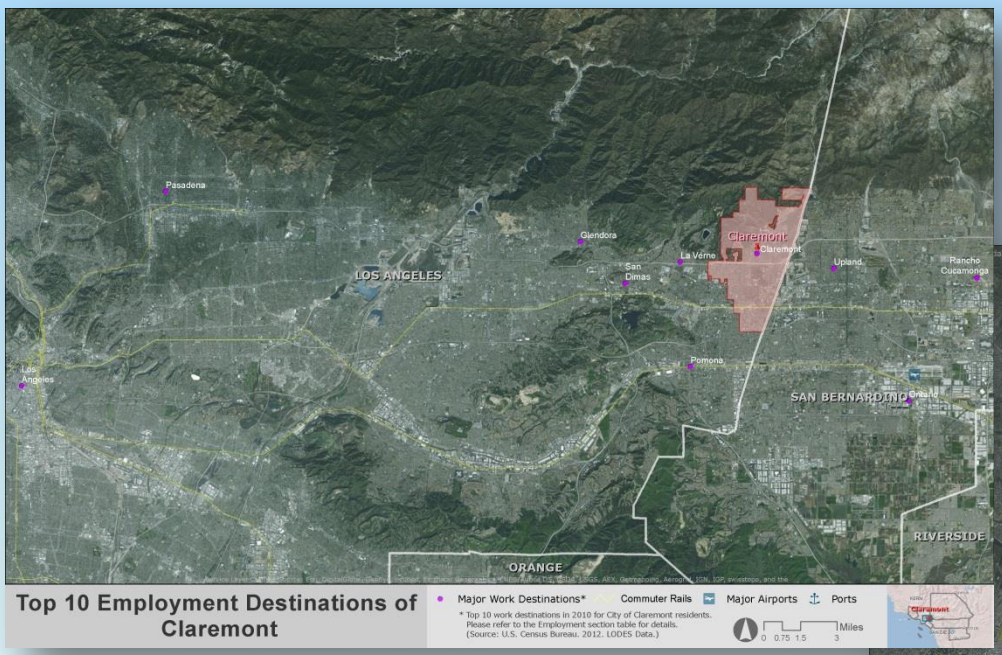
Python Script to Visualize Top 10 Work Destinations

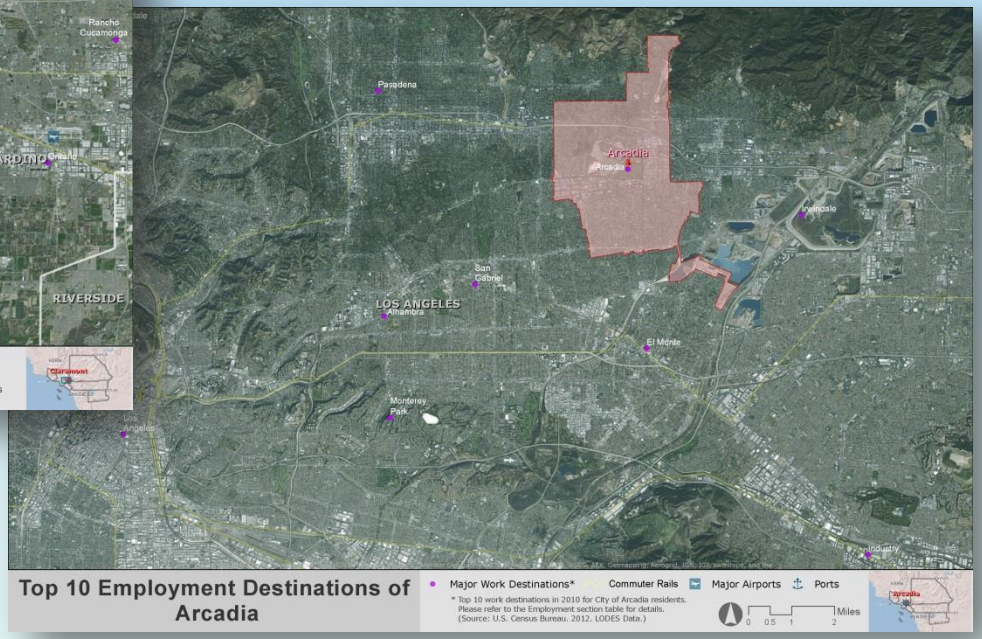
- Importing ArcPy and ArcPy.mapping modules in Python to automatically create work destination maps
  - Data Driven Page, Definition Query, and Exporting functions in ArcGIS
    - mxd.dataDrivenPages.currentPageID
    - lyr.definitionQuery = sqlExp
    - arcpy.mapping.ExportToJPEG

```
# Import ArcPy and os modules
import arcpy, os
import arcpy.mapping

#Set the current workspace
arcpy.env.workspace = "P:/LEHD/esri_uc_2014/shapes/local_profiles_2013.gdb"

#Set the current map document
mxd = arcpy.mapping.MapDocument("P:/LEHD/esri_uc_2014/mxds/regional_maps/regional_map_city_gnrl.mxd")

# Define data driven page variables
totalPage = mxd.dataDrivenPages.pageCount
curPage = mxd.dataDrivenPages.currentPageID
```

Visualization of Top 10 Work Destinations

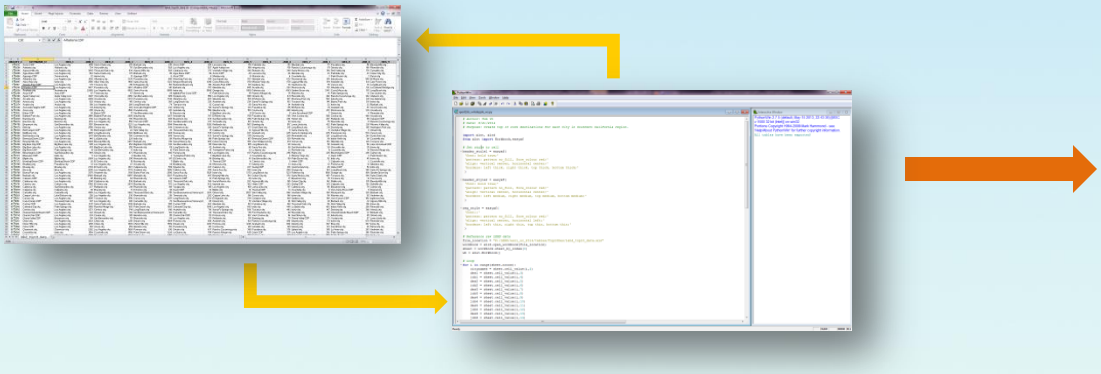
- Created a series of maps that depict top 10 work destinations for 191 cities and 6 counties





Incorporation of Python and MS Excel

- Create Top 10 Destination tables for each jurisdiction
  - E.g. xlrd and xlwt module
    - citynames = sheet.cell_value(i,2)
    - s.write(0,0,citynames,header_style1)



Alhambra city	
Cities	**Job #**
Los Angeles city	9812
Alhambra city	3335
Pasadena city	1691
Monterey Park city	1115
San Gabriel city	767
Glendale city	672
Rosemead city	614
Industry city	554
El Monte city	534
Burbank city	513
Total	**19607**

Conclusions

- Significantly efficient in processing O-D analysis with SAS for numerous jurisdictions
- Some inconsistencies issue in the total number of job at place level between SAS programming *and OnTheMap* application
- Data limitation of *LODES* datasets
- Python and ArcGIS improve the processing time and accuracy of map production

Future Studies

- Work destinations at different geographical level
  - City to City → Sub-region to Sub-region
  - Top 10 work destinations → Top 20 work destinations
- Produce major work destination data using Census Transportation Planning Products (CTPP)
- Develop Desktop application or ArcGIS toolbox
- Web-based mapping to identify major work destinations for multiple jurisdictions

Thank you!

Yunkyung Choi (ChoiY@scag.ca.gov)

Tom M. Vo (Vo@scag.ca.gov)

SOUTHERN CALIFORNIA ASSOCIATION of GOVERNMENTS